Face aux métadonnées

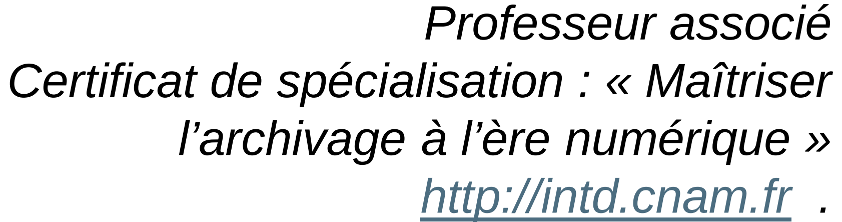
Carte d’orientation dans la jungle des métadonnées…

Groupe PIN
21 juin 2011

Marie-Anne Chabin
marie-anne.chabin@archive17.fr
www.archive17.fr



Marie-Anne Chabin



Expert en Archivage / Records management,
www.archive17.fr

Professeur associé
Certificat de spécialisation : « Maîtriser l'archivage à l'ère numérique »
http://intd.cnam.fr .



Chercheur, auteur, conférencier

- *Nouveau glossaire de l'archivage (en ligne sur www.archive17fr)*
- *Méthode Arcateg™*



Secrétaire générale du CR2PA
www.cr2pa.fr



Archiviste départemental
Conservateur en chef à la DAF
(1984-1995)





Responsable de la vidéothèque d'actualité de l'INA
(1997-1999)

Démarche suivie

1. Collecte
   - définitions (*normes et blogs*), exemples, modèles…
2. Analyse
   - dissection, classement, comparaison
3. Reconstruction
   - hiérarchie, cartographie

<u>Concept de base</u>:
Un objet d'information / un document =
un ensemble d'informations fini (daté)
ayant du sens pour son auteur
utilisable, réutilisable, révisable

Définition et catégorisation des métadonnées

I – Définitions etc.

- Diverses définitions
- Des comparaisons parlantes
- Origine, finalités, évolutions
- Les métadonnées avant les métadonnées
- Exemples

Précisions linguistiques

- *Data* : données (objet d'information)
  - pas de sens au singulier
- *Metadata* : Métadonnées
- *Metadata element* : métadonnée
  - (l'anglais n' a pas de singulier pour *data* et *metadata*)

- Cousins et voisins
  - attributs
  - index
  - annotations
  - information de représentation
  - propriétés
  - tag

propriétés et valeurs

Origine (1)

- Au début était **le web** et sa croissance exponentielle, et il fallait y être, on s'est alors intéressé au référencement, au développement des annuaires et des moteurs de recherches.
- Et puis la grenouille devenant aussi grosse que le bœuf… C'est cette idée qui a accompagné les premières réflexions sur "l'organisation" du web.
- Dans un premier temps, il y a eu les métadonnées HTML, celles des **balises** librement complétées ou non par les auteurs de site web.

http://aristide.12.free.fr/spip.php?article142

Origine (2)

- Les métadonnées décrivent différents attributs d'informations et leur donnent signification, contexte et organisation. La pratique et la théorie des métadonnées est familière à certains, ses origines remontant au **catalogage** (indexation) de publications écrites. Dans le monde numérique, des catégories supplémentaires de métadonnées sont apparues afin de gérer la navigation et la gestion de fichiers.

http://www.library.cornell.edu/preservation/tutorial-french/metadata/metadata-01.html

Définition (1)

- Le terme barbare de « **métadonnées** » désigne tout ce qui va « **enrichir** » votre livre électronique et lui ajouter des informations supplémentaires qui permettront au lecteur d'en consulter un résumé, de savoir à quel(s) genre(s) il appartient, de connaître son numéro d'enregistrement (ISBN), son prix, etc.

http://blog.tcrouzet.com/2010/10/04/texte-vers-epub/

Définition (2)

- Une métadonnée (mot composé du préfixe grec meta, indiquant l'auto-référence ; le mot signifie donc proprement « donnée de/à propos de donnée ») est une donnée servant à **définir ou décrire** une autre donnée quel que soit son support (papier ou électronique).
- Les métadonnées sont à la base des techniques du Web sémantique. Elles sont définies dans le cadre du modèle Resource Description Framework (RDF).

Wikipédia

Définition (3)

- Les métadonnées, c'est-à-dire une donnée servant à **définir ou décrire** une autre donnée, elles permettent la localisation et le référencement des données présentes dans les systèmes de gestion de base de données des différents acteurs de l’aménagement du territoire.
- Les métadonnées permettent donc à n’importe quels utilisateurs de savoir :
  - - Quelles données existent ?
  - - La pertinence au regard d’une problématique
  - - Comment y avoir accès ?
- Les métadonnées ont un rôle d’identification et de spécification : qui, quand, quoi, où?

http://seine-aval.crihan.fr/web/attached_file/componentId/kmelia232/attachmentId/17814/lang/fr/name/rapport_D_Lemaire.pdf

Définition (4)

- Les métadonnées sont un ensemble de données structurées **décrivant** des ressources physiques ou numériques. Elles sont un maillon essentiel pour le **partage** de l'information et **l'interopérabilité** des ressources électroniques.
- Elles sont classiquement divisées en métadonnées descriptives, administratives ou de structure.

http://openaccess.inist.fr/spip.php?mot3

Définition (5)

- Le terme « métadonnées » définit l'ensemble des informations utilisées pour décrire d'autres données. Une très brève définition des métadonnées est alors «données **concernant** des données ». Les descriptions des métadonnées vont au-delà de la forme et du contenu purs des données pour englober les **faits administratifs** concernant les données (l'auteur et la date de leur création) ainsi que leur mode de collecte et de traitement avant d'être diffusées ou archivées dans une base de données. En outre, les métadonnées permettent une recherche efficace de données et de leur localisation.

http://unstats.un.org/unsd/EconStatKB/Attachment384.aspx
Recommandations internationales concernant les statistiques industrielles, 2008 - 8.15.

Définition (6)

- Une première définition des métadonnées a été fournie au début de l'ère des banques de données, dans les années septante : « les métadonnées sont des données sur les données ». Les systèmes d'information n'ont cessé de gagner en complexité ...
- Ainsi, une définition plus actuelle et plus générale est formulée comme suit : « **les métadonnées constituent des déclarations émises par un niveau d'abstraction supérieur concernant un niveau inférieur** ».
- Les métadonnées sont une documentation sur les données, prioritairement destinée à servir d'information, permettant à l'utilisateur ou au gestionnaire de comprendre, de comparer et d'échanger le contenu du jeu de données décrit.

Modèle de métadonnées GM03
http://www.geocat.ch/internet/geocat/fr/home/documentation/gm03.parsys.5359.downloadList.69384.DownloadFile.tmp/gm03v23fr.pdf

Définition (7)

- Les métadonnées sont des données sur les données
- En ce sens, les métadonnées sont **une représentation d'une représentation de la réalité**.
- Les métadonnées sont utiles sur plusieurs plans :
  - « availability -- data needed to determine the sets of data that exist for a geographic location.
  - fitness for use -- data needed to determine if a set of data meets a specific need.
  - access -- data needed to acquire an identified set of data.
  - transfer -- data needed to process and use a set of data »

http://www.fgdc.gov/metadata/csdgm/

Définition (8)

- Les métadonnées sont les informations **associées** aux documents gérés.
Elles servent à qualifier les documents, leur apporter une information complémentaire qui soit exploitable. La métadonnée est directement utile en tant qu'information, mais surtout elle est la base des fonctions de recherche, ou de **traitements sélectifs** appliqués aux documents.

http://ged-open-source.smile.fr/Enjeux-principaux-de-la-GED/Structuration-des-metadonnee

Comparaison (1)

- **EXIF** est l'abréviation d'*Exchangeable image file format*.

  Les données EXIF constituent un remplacement commode du **petit carnet** qui accompagnait à l'époque de la photographie chimique les photographes méticuleux. Sur les appareils numériques, elles sont bien entendu conservées automatiquement avec chaque photo.

Wikipédia

Comparaison (2)

- « les métadonnées sont aux données ce que l'étiquette est à un médicament »
  - renseignements comme le nom du médicament, celui du fabricant, la posologie, la composition chimique, les précautions à prendre, le nombre de comprimés et la date de péremption du médicament
  - favoriser l'utilisation adéquate du médicament ; informer les usagers de leur existence, de leurs caractéristiques et des précautions d'emploi à respecter, effets secondaires



(vocabulaire de la géomatique, IGN)

Jean Denègre, EN Sciences Géographiques

Comparaison (3)

- Votre fichier va se balader tout au long de sa carrière de fichier avec une **petite valise**. Dans cette petite valise on va pouvoir y trouver son journal intime et son carnet de naissance. Sur son carnet de naissance, il y aura le nom de son père (ou de sa mère) là où il est né, et quand. Dans le journal, chaque personne que le fichier va rencontrer va marquer qui il est, à quelle heure ils se sont vus, où et ce qu'ils ont fait.

http://www.labo-security.com/2011/03/06/metadonnees-ou-les-donnees-vicieuses/

Exemple (1)

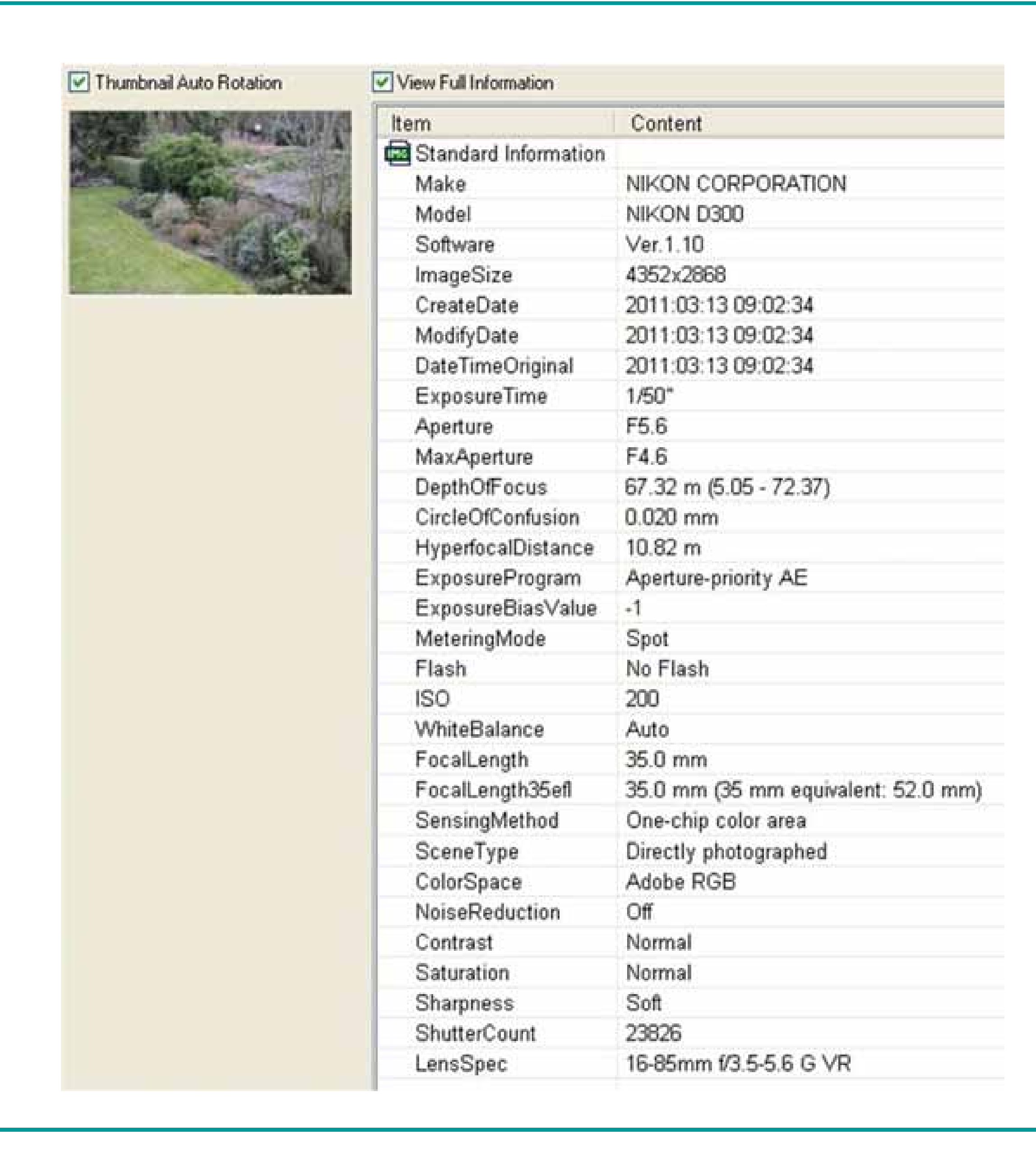
Métawal - Information géographique wallonne





N° ISO	Attribut	Dénomination française	Statut	Occurrence	Type de données / Format	Valeur par défaut	Domaine
1 Metadata entity set information (MD_Metadata)							
1	MD_Metadata	Informations sur les métadonnées	M	1	Classe		
2	fileIdentifier	Identifiant	M	1	Chaîne de caractères		Texte libre
3	language	Langue	M	1	Classe (Liste déroulante)	fr	MD_LanguageCode
4	characterSet	Jeu de caractères	M	1	Classe (Liste déroulante)	UTF-8	MD_CharactersetCode
5	parentIdentifier	Identifiant du fichier parent	C	1	Chaîne de caractères		Texte libre
6	hierarchyLevel	Type de ressources auxquelles les métadonnées s'appliquent	M	N	Classe (Liste déroulante)	dataset	MD_ScopeCode
7	hierarchyLevelName	Nom du type de ressources auxquelles les métadonnées s'appliquent					
8	contact	Organisme/Personne responsable	M	N	Classe (Liste à choix multiple)		CI_ResponsibleParty
9	dateStamp	Dernière date de mise à jour	M	1	Date (JJ/MM/AAAA)	Date du jour	Date (Mise à jour à chaque modification de la métadonnée)
10	metadataStandardName	Nom de la norme	M	1	Chaîne de caractères	ISO 19115 - Profil wallon	Texte libre
11	metadataStandardVersion	Version de la norme	M	1	Chaîne de caractères	FDIS - version 1	Texte libre
11.1	dataSetURI	URI du jeu de données					
12	spatialRepresentationInfo	Représentation spatiale du jeu de données	O	N	Classe (Liste à choix multiple)		MD_SpatialRepresentation
13	referenceSystemInfo	Système de référence spatiale du jeu de données	O	N	Classe (Liste à choix multiple)		MD_ReferenceSystem
14	metadataExtensionInfo	Extensions de la norme	O	N	Classe (Liste à choix multiple)		MD_MetadataExtension
15	identificationInfo	Identification du jeu de données	M	N	Classe (Liste à choix multiple)		MD_Identification
16	contentInfo	Contenu du jeu de données	O	N	Classe (Liste à choix multiple)		MD_ContentInformation
17	distributionInfo	Modalités de distribution du jeu de données	O	1	Classe (Liste déroulante)		MD_Distribution
18	dataQualityInfo	Qualité du jeu de données	O	N	Classe (Liste à choix multiple)		DQ_DataQuality
19	portrayalCatalogueInfo	Catalogue des règles de représentation du jeu de données					
20	metadataConstraints	Contraintes sur les métadonnées					
21	applicationSchemaInfo	Schéma conceptuel du jeu de données					
22	metadataMaintenance	Mise à jour					
22W3	attributetableInfo	Tables attributaires du jeu de données	O	N	Classe (Liste à choix multiple)		AD_TableIdentification

Exemple (2)



Thumbnail Auto Rotation

View Full Information

Item	Content
Standard Information	
Make	NIKON CORPORATION
Model	NIKON D300
Software	Ver.1.10
ImageSize	4352x2868
CreateDate	2011:03:13 09:02:34
ModifyDate	2011:03:13 09:02:34
DateTimeOriginal	2011:03:13 09:02:34
ExposureTime	1/50"
Aperture	F5.6
MaxAperture	F4.6
DepthOfFocus	67.32 m (5.05 - 72.37)
CircleOfConfusion	0.020 mm
HyperfocalDistance	10.82 m
ExposureProgram	Aperture-priority AE
ExposureBiasValue	-1
MeteringMode	Spot
Flash	No Flash
ISO	200
WhiteBalance	Auto
FocalLength	35.0 mm
FocalLength35efl	35.0 mm (35 mm equivalent: 52.0 mm)
SensingMethod	One-chip color area
SceneType	Directly photographed
ColorSpace	Adobe RGB
NoiseReduction	Off
Contrast	Normal
Saturation	Normal
Sharpness	Soft
ShutterCount	23826
LensSpec	16-85mm f/3.5-5.6 G VR

Exemple (3)

Métadonnées des référentiels de l'Herbier de l'Université de Strasbourg et des bases de données « BRUNFELS » et Herbier de Strasbourg, 2010

Structure des données du fichier « BRUNFELS »

NOTE : (*) champ non diffusé systématiquement lors des échanges

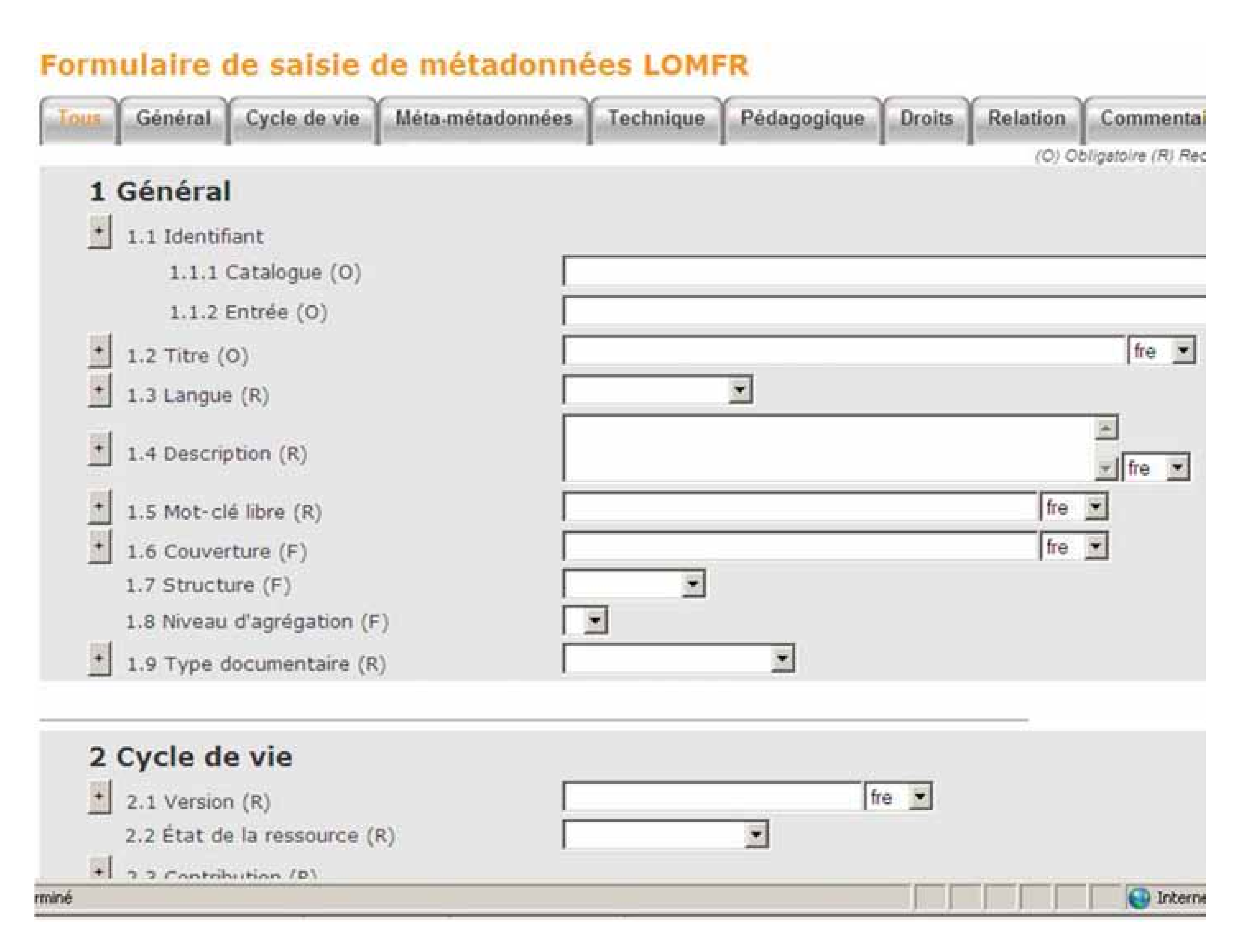
NU_BIBLIO : numéro de la référence bibliographique (la référentiel bibliographique est disponible sous simple demande).
ORIG : origine institutionnelle de la donnée (annexe 1)
NU_COL : numéro de référence de l'observateur ou de la source
(*) BOTANISTE : nom du (ou des) botaniste(s) indiqué dans la source
NU_BDNFF : numéro du taxon dans la banque de données nationale de la Flore de France (BDNFF v4.02). Site Tela-Botanica : http://www.tela-botanica.org/
NU_CIFF : numéro du taxon dans le Code Informatisé de la Flore de France (Brisse et

INTRODUIT : plante introduite (non indigène), l'introduction pouvant être volontaire - plante alimentaire ou ornementale cultivée, ou fortuite - adventice.
CULTIVE : plante cultivée : alimentaire, ornementale, forestière, tinctoriale, etc.
ENVAHISSANTE : plante envahissante ou potentiellement envahissante
MESICOLES : taxon du Plan national d'Action Messicoles
NA2000 : commune située en tout ou partie dans un site Natura 2000 (annexe 3)

Métadonnées des bases de données « BRUNFELS » et Herbier de Strasbourg

NU_BIBLIO : numéro de la référence bibliographique (la référentiel bibliographique est disponible sous simple demande).
ORIG : origine institutionnelle de la donnée (annexe 1)
NU_COL : numéro de référence de l'observateur ou de la source
(*) BOTANISTE : nom du (ou des) botaniste(s) indiqué dans la source
NU_BDNFF : numéro du taxon dans la banque de données nationale de la Flore de France (BDNFF v4.02). Site Tela-Botanica : http://www.tela-botanica.org/
NU_CIFF : numéro du taxon dans le Code Informatisé de la Flore de France (Brisse et Kerguélen 1993) révisé M. Hoff (2010) pour le Rhin supérieur
(*) TAXON : nom du taxon indiqué dans la source
JOUR de l'observation, s'il y a lieu
MOIS de l'observation s'il y a lieu
ANNEE de l'observation ou de la publication de l'information
DATEXACT : * si la date est effectivement celle de l'observation
ALTITUDE : s'il y a lieu

Exemple (4)



Formulaire de saisie de métadonnées LOMFR
Tous
Général
Cycle de vie
Méta-métadonnées
Technique
Pédagogique
Droits
Relation
Commentai
(O) Obligatoire (R) Rec
1 Général
1.1 Identifiant
1.1.1 Catalogue (O)
1.1.2 Entrée (O)
1.2 Titre (O)
fre
1.3 Langue (R)
1.4 Description (R)
fre
1.5 Mot-clé libre (R)
fre
1.6 Couverture (F)
fre
1.7 Structure (F)
1.8 Niveau d'agrégation (F)
1.9 Type documentaire (R)
2 Cycle de vie
2.1 Version (R)
fre
2.2 État de la ressource (R)
rminé
Interne

Exemple (5)

Saisie des métadonnées d'une donnée géographique

Description

*Titre de la donnée INSPIRE :

Mots-clés décrivant les données

Le formulaire permet de remplir les métadonnées décrivant une donnée géographique selon la norme internationale ISO 19115 (Métadonnées).
Ce formulaire indique uniquement les métadonnées "minimales" nécessaires pour inclure la description de ces données dans le Géocatalogue (www.geocatalogue.fr). Il est conseillé de compléter par des informations proposées par la norme permettant de constituer une vue réaliste et détaillée des données cartographiques que vous disposez.
L'asterisque rouge indique les informations à remplir obligatoirement.

Informations sur la métadonnée

*Identifiant de la métadonnée :

Langue décrivant les métadonnées INSPIRE :

géoportail le portail des territoires & des citoyens

Exemple (6)

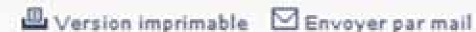
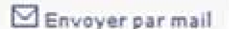
Nuages de mots



Commentaire (1)

Publié le lundi 4 avril 2011

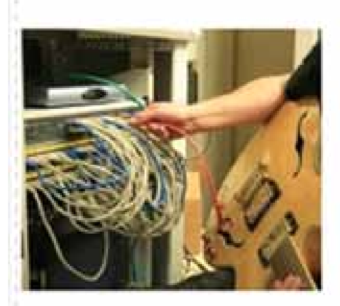
Article

Version imprimable   Envoyer par mail

L'ENJEU DES MÉTADONNÉES MUSICALES
La musique n'est pas qu'un fichier son

Payante ou gratuite, la musique circule plus que jamais sur le flux numérique ou en se fixant sur les disques durs. Mais la dématérialisation lui a fait perdre ses données associées. Généralement, on a le nom du morceau et celui de l'interprète, mais rarement plus. Sans parler du besoin d'identification pour les ayants droit. C'est tout l'enjeu des métadonnées pour la musique et c'est le thème d'une live-conférence organisée à la Mix Box le jeudi 28 avril [voir encadré].



Autrefois, nombre d'informations utiles apparaissaient dans les livrets des CD qui comportaient souvent des photos, éléments de biographie, paroles de chanson, etc. Paradoxalement, le numérique, qui pourrait permettre d'accéder à encore plus d'informations sur les fichiers musicaux — ce qu'on désigne donc comme métadonnées — a souvent tendance à appauvrir leur mise en contexte.

Métadonnées

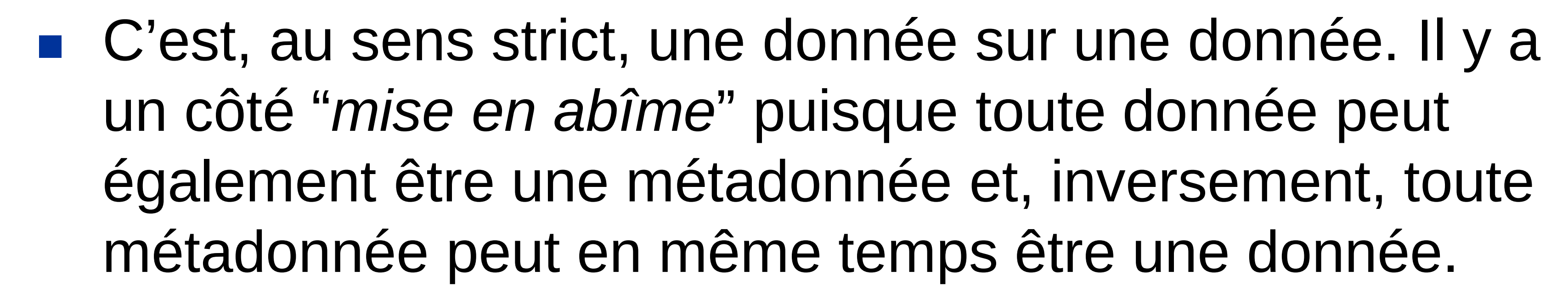
Une première raison d'être des métadonnées pour la musique est l'identification des œuvres et de tous ceux qui ont contribué à sa création

http://www.irma.asso.fr/L-ENJEU-DES-METADONNEES-MUSICALES

Commentaire (2)



- Une métadonnée est aussi une donnée.



- C'est, au sens strict, une donnée sur une donnée. Il y a un côté "*mise en abîme*" puisque toute donnée peut également être une métadonnée et, inversement, toute métadonnée peut en même temps être une donnée.
- Un système d'indexation *full text* est donc un système technologique qui fait de chaque mot une métadonnée, faisant ainsi l'économie du recourt à un système de classement, c'est à dire un système de métadonnées, dont il faudrait connaître le fonctionnement *a priori*.

http://www.christian-faure.net/2011/03/09/sur-les-metadonnees/

Commentaire (3)

- … Dans le monde physique, on peut facilement faire la différence entre le livre et la carte du catalogue qui en parle. En ligne, il n'y a plus de distinction entre données et métadonnées. […], **tout est métadonnées. La seule différence est que les données sont ce que vous cherchez et les métadonnées ce que vous savez**.

Comment le web change le monde. L'alchimie des multitudes. http://alchimiedesmultitudes.atelier.fr
Francis Pisani & Dominique Piotet

Les métadonnées avant les métadonnées (1)



Manuscrit du 13e siècle
avec gloses et annotations
sites.google.com



Registre des interventions chirurgicales
ou obstétricales
Berger Levrault Ref. 530603

Les métadonnées avant les métadonnées (2)

R	BALZAC (Honoré de).-
BAL CES	[Histoire de la grandeur et de la décadence de] César Birotteau, parfumeur, Chevalier de la Légion d'Honneur, Adjoint au Maire du Deuxième Arrondissement de la Ville de Paris.- Paris, A.Michel, [1954] .- 19 cm., 423 p., 70800 (Scènes de la vie parisienne.)

Fiche de catalogage dactylographiée

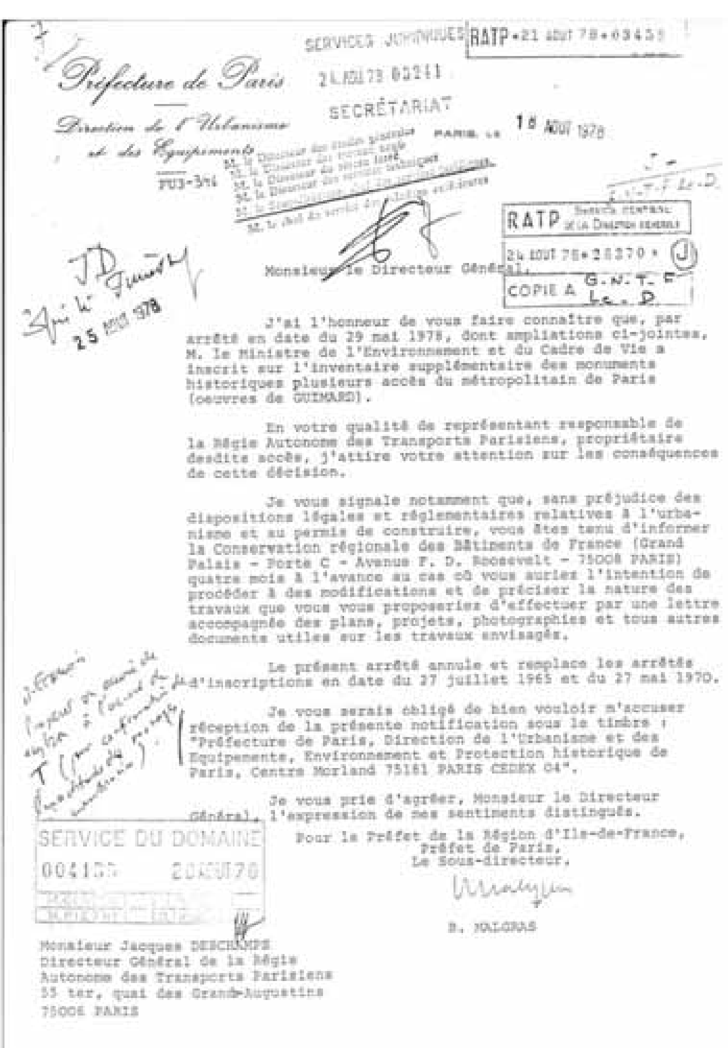
458. (Don Charles Gomart). « Charte (présumée) de la commune de Ham, en 1188 *(sic)*, autographiée par Ch. Gomart en 1863.d'après l'original qui se trouve aux archives de la ville d'Albert. » 1178 copie.

459. (Don de La Hautoie. cf. rapport 1899, p. 290-291). Notes historiques et copie de pièces sur l'abbaye de St-Prix de St-Quentin (18e-19e siècles). Copie de pièces depuis 986. 986 copie-19e siècle.

Archives départementales de la Somme Répertoire numérique de la série F

Les métadonnées avant les métadonnées (3)

Correspondance administrative (1978)



Préfecture de Paris

SECRÉTARIAT

RATP • 21 AOUT 78 • 03435

Direction de l'Urbanisme et des Equipements

PU3-346

PARIS, LE 18 AOUT 1978

RATP SERVICE CENTRAL DE LA DIRECTION GENERALE

24 AOUT 78 • 28370 •

COPIE A G.N.T.F. Le D.

Monsieur le Directeur Général,

J'ai l'honneur de vous faire connaître que, par arrêté en date du 29 mai 1978, dont ampliations ci-jointes, M. le Ministre de l'Environnement et du Cadre de Vie a inscrit sur l'inventaire supplémentaire des monuments historiques plusieurs accès du métropolitain de Paris (oeuvres de GUIMARD).

En votre qualité de représentant responsable de la Régie Autonome des Transports Parisiens, propriétaire desdits accès, j'attire votre attention sur les conséquences de cette décision.

Je vous signale notamment que, sans préjudice des dispositions légales et réglementaires relatives à l'urbanisme et au permis de construire, vous êtes tenu d'informer la Conservation régionale des Bâtiments de France (Grand Palais - Porte C - Avenue F. D. Roosevelt - 75008 PARIS) quatre mois à l'avance au cas où vous auriez l'intention de procéder à des modifications et de préciser la nature des travaux que vous vous proposeriez d'effectuer par une lettre accompagnée des plans, projets, photographies et tous autres documents utiles sur les travaux envisagés.

Le présent arrêté annule et remplace les arrêtés d'inscriptions en date du 27 juillet 1965 et du 27 mai 1970.

Je vous serais obligé de bien vouloir m'accuser réception de la présente notification sous le timbre : "Préfecture de Paris, Direction de l'Urbanisme et des Equipements, Environnement et Protection historique de Paris, Centre Morland 75181 PARIS CEDEX 04".

Je vous prie d'agréer, Monsieur le Directeur Général, l'expression de mes sentiments distingués.

Pour le Préfet de la Région d'Ile-de-France,
Préfet de Paris,
Le Sous-directeur,

B. MALGRAS

SERVICE DU DOMAINE

Monsieur Jacques DESCHAMPS
Directeur Général de la Régie
Autonome des Transports Parisiens
53 ter, quai des Grands-Augustins
75006 PARIS

II – Catégorisation, organisation

- Typologies
- Normes, schémas, profils, registres de métadonnées
- Normes relatives à la gestion de l'information incluant un modèle de métadonnées

Typologie (1)

- Il ne faut pas réduire les métadonnées d'un document à sa description bibliographique. A côté des **métadonnées descriptives**, il existe aussi des **métadonnées techniques** (format, techniques de production, support...), des **métadonnées administratives** (cycle de vie d'un document dans son contexte de production, d'utilisation et de conservation), des métadonnées juridiques ou encore des métadonnées structurelles (structure logique ou physique d'un document).
- Dans TEF, ces différents types de métadonnées sont appelés "métadonnées de gestion". TEF porte sur les métadonnées descriptives et sur les métadonnées de gestion des thèses électroniques.

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

Typologie (2)

Cinq objectifs pour les métadonnées

- Faciliter la **recherche** d'information
  - décrire le contenu et les relations entre les fichiers d'un site,
  - classer le contenu suivant un degré de difficulté ou un public cible,
  - mieux référencer un site ou une page sur Internet.
- Faciliter **l'interopérabilité**
  - partager et échanger des informations.
- Faciliter la **gestion et l'archivage**
  - informer sur le cycle de vie des documents,
  - gérer des collections de ressources,
  - gérer des archives électroniques.
- Gérer et protéger les **droits**
  - les droits de propriété intellectuelle,
  - les droits d'accès à des pages web (restrictions de consultation).
- **Authentifier** un texte
  - encoder une signature électronique pour valider un texte sur Internet

http://www.educnet.education.fr/dossier/metadata/@@document_whole2

Typologie (3)

	thèse	version	édition	fichier	ressource externe
MD descriptives	tef_desc_these	tef_desc_version	tef_desc_edition		tef_desc_externe
MD administratives	tef_admin_these				
MD de droits	tef_droits_etablissement_these tef_droits_auteur_these	tef_droits_version			tef_droits_externe
MD de conservation				tef_tech_fichier	

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

Typologie (4)

- Pour des raisons pratiques, les types et les fonctions de métadonnées peuvent être classés en trois catégories générales :
  - descriptif,
  - structurel et
  - administratif.
- Ces catégories n'ont pas toujours de frontières parfaitement définies et **se chevauchent parfois**. Par exemple, les métadonnées administratives contiennent une grande gamme d'informations pouvant être considérées comme des métadonnées descriptives et structurelles

http://www.library.cornell.edu/preservation/tutorial-french/metadata/metadata-01.html

Typologie (5)

TYPE	BUTS	EXEMPLES D'ELEMENTS
Métadonnées Descriptives	Décrivent et identifient les ressources d'information au niveau local (système) afin d'autoriser la recherche et l'extraction (p.ex., chercher des peintures d'animaux) au niveau du web, en permettant aux usagers de découvrir les ressources (p.ex. cherchez des collections numériques de poésie)	Identifiants uniques (PURL, Handle) Attributs physiques (médias, dimensions, état général) Attributs bibliographiques (titre, auteur/créateur, langue, mots-clés)
Métadonnées structurelles	Facilitent la navigation et la présentation des ressources électroniques Fournissent des informations sur la structure interne des ressources y compris la page, section, le numéro de chapitre, les index, et la table des matières Décrivent les relations entre les matériels (p.ex., la photographie B a été insérée dans le manuscrit A) Relient les fichiers et scripts liés (p. ex. Le fichier A est le format JPEG du fichier d'archive B)	Balises structurantes telles que titre de page, table des matières, chapitres, parties, erratum, index, relation sous-objet (p.ex. photographie d'un agenda)
Métadonnées Administratives	Facilitent la gestion et le traitement à court et long termes des collections numériques Incluent des données techniques sur la création et le contrôle de la qualité incluent la gestion des droits, le contrôle d'accès et les conditions utilisateur requises conservation action information	Données techniques telles que Type et modèle de scanner, résolution, profondeur de bit, espace colorimétrique, format de fichier, compression, source lumineuse, propriétaire, date du copyright, restrictions de copie et de distribution, informations de licence, activités de conservation (cycles de rafraîchissement, migration, etc…)

http://www.library.cornell.edu/preservation/tutorial-french/metadata/metadata-01.html

Origines des normes (1)

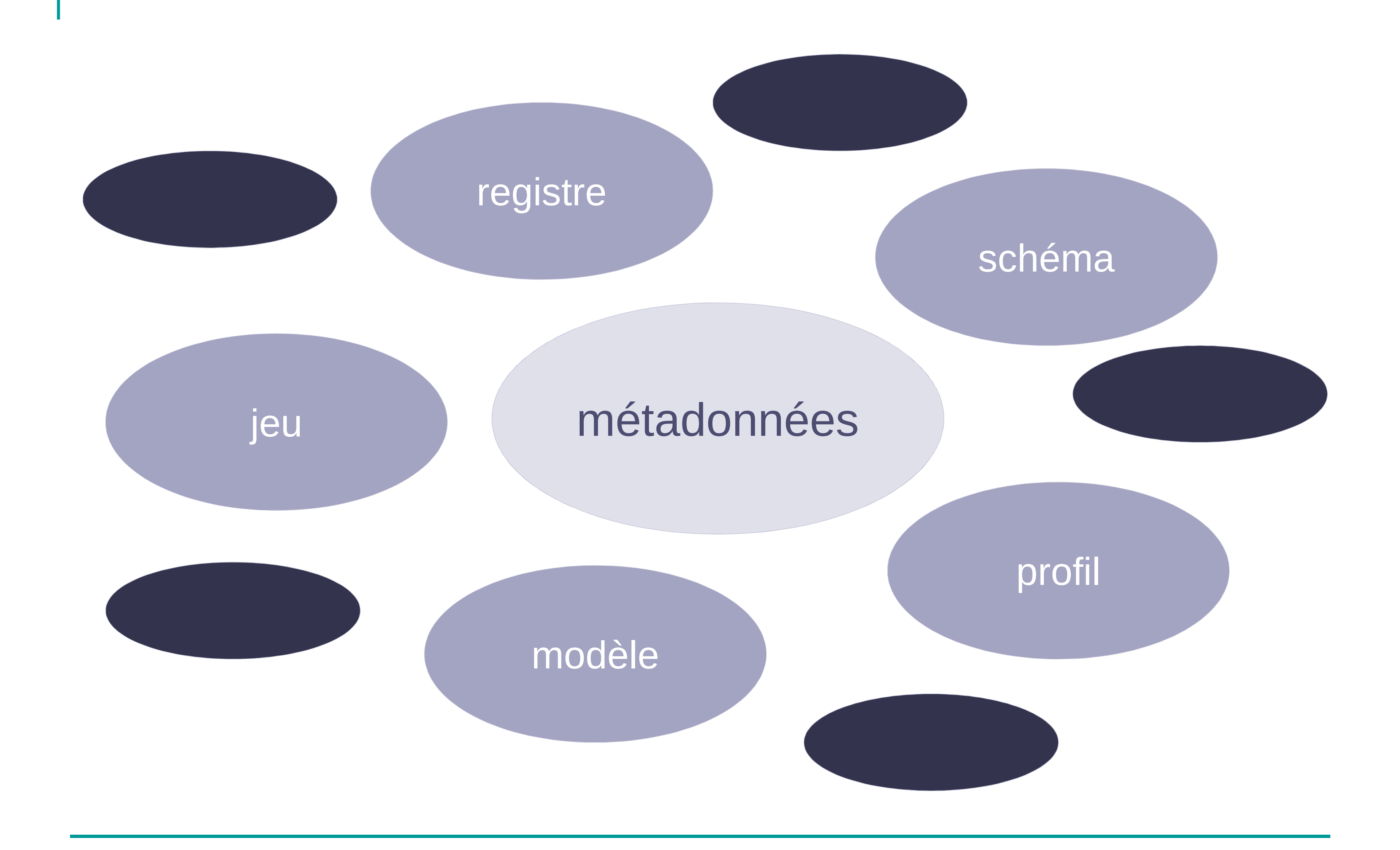
- Les métadonnées sont structurées selon une **hiérarchie dans l'importance des questions à laquelle elles doivent répondre**, il est nécessaire de se **baser sur une norme** sachant que l'utilisation d'un standard est un gain de temps et de qualité.

http://seine-aval.crihan.fr/web/attached_file/componentId/kmelia232/attachmentId/17814/lang/fr/name/rapport_D_Lemaire.pdf

Origines des normes (2)

- … Puis des professionnels de l'information ont commencé à réfléchir à cette organisation. Ainsi en 1995, à l'initiative du gouvernement américain un groupe de travail s'est constitué et a abouti à la mise en place du Dublin Core un schéma de métadonnées standard.
- Ce sont ensuite développés le LOM (Learning Object Metadata) orienté vers l'enseignement et son adaptation française LOMFR.
- Ces schémas facilitent la compréhension des données pour les hommes et les machines, et permettent mutualisations et échanges comme les OAI (Open Archives intiative).

http://aristide.12.free.fr/spip.php?article142


registre
schéma
jeu
métadonnées
profil
modèle

Registre de métadonnées

- Un registre de métadonnées est, selon la définition qu'en donne le Dublin Core dans l'ébauche finale du 24 février 2001, un « **Système de gestion** des métadonnées, c'est-à-dire un système formel qui fournit l'information d'autorité sur la sémantique et la structure de chaque élément »

- Les registres de métadonnées sont utilisés à chaque fois que des données doivent être utilisées de façon cohérente **dans une organisation** ou un groupe d'organisations :
  - Administration et ses établissements déconcentrés,
  - Informations géographiques,
  - Pôles de compétitivité,

http://fr.wikipedia.org/wiki/Registre_de_m%C3%A9tadonn%C3%A9es

Schéma d'encodage

- **Référentiel** qui aide à l'interprétation de la valeur d'un élément Dublin Core.
- Un schéma d'encodage peut être un **vocabulaire** contrôlé ou une notation formelle qui indique la manière correcte d'exprimer une valeur. Dans TEF, par exemple, le vocabulaire Rameau et la norme 639-1 (codes de langue) sont utilisés comme schéma d'encodage.
- Voir aussi Dublin Core qualifié, **Qualificatif de raffinement**.

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

Dublin Core



- Dublin Core qualifié (*Dublin Core Metadata Terms*)
- **Ensemble** de métadonnées regroupant les 15 éléments du Dublin Core simple ainsi que d'autres termes qui permettent de compléter et de préciser ceux-ci. Parmi ces autres termes, on trouve surtout les deux types de "qualificatifs" : **les qualificatifs de raffinement et les schémas d'encodage.**

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

Les 15 méta-données du Dublin Core

Nom de l'élément	Identifiant	Définition
Titre	Title	Le nom donné à la ressource
créateur	Creator	L'entité principalement responsable de la création du contenu de la ressource
sujet et mots-clefs	Subject	Le sujet du contenu de la ressource
description	Description	Une description du contenu de la ressource
éditeur	Publisher	L'entité responsable de la diffusion de la ressource, dans sa forme actuelle, tels, un département universitaire, une entreprise.
contributeur	Contributor	Une entité qui a contribué à la création du contenu de la ressource
date	Date	Une date associée avec un événement dans le cycle de vie de la ressource
Type	Type	La nature ou le genre du contenu de la ressource
Format	Format	La matérialisation physique ou digitale de la ressource
Identifiant	Identifier	Une référence non ambiguë à la ressource dans un contexte donné
source	Source	Une référence à une ressource à partir de laquelle la ressource actuelle a été dérivée
Langue	Language	La langue du contenu intellectuel de la ressource
Relation	Relation	Une référence à une autre ressource qui a un rapport avec cette ressource
couverture	Coverage	La portée ou la couverture spatio-temporelle de la ressource
droits	Rights	Information sur les droits sur et au sujet de la ressource

cairn.info

RDF

RDF (Resource Description Framework)

- **Modèle** organisant les métadonnées sous la forme d'assertions élémentaires composées d'un sujet, d'un prédicat et d'un objet.
- Un triplet RDF est une association :(sujet, prédicat, objet)
  - Le sujet représente la ressource à décrire ;
  - Le prédicat représente un type de propriété applicable à cette ressource ;
  - L'objet représente une donnée ou une autre ressource : c'est la valeur de la propriété.
- http://www.w3.org/RDF/.

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

ISO 19115 et ISO 19139

- L'ISO 19115-2:2009 étend la norme existante sur les métadonnées géographiques en définissant le schéma requis de description des images et des matrices. Elle donne des informations relatives aux propriétés des équipements de mesure utilisés pour obtenir les données, la géométrie du processus de mesure utilisée par ces équipements, et le processus de production utilisé pour numériser les données brutes.

http://www.iso.org

Pour l'échange des métadonnées, la norme ISO 19139 traduit la norme ISO 19115 au format XML (eXtensible Markup Language) pour permettre une **standardisation et l'interopérabilité** des systèmes en vue d'échanger et mutualiser les métadonnées

PREMIS

- **Preservation Metadata : Implementation Strategies)**
- **dictionnaire** de données définissant des éléments principaux afin d'améliorer les fonctions de conservation. Il liste les métadonnées de conservation qui doivent être connues du service d'archives.

www.cines.fr

METS (1)

- METS (*Metadata Encoding and Transmission Standard*)
- METS est un **standard** maintenu par la Bibliothèque du Congrès. C'est un **vocabulaire** XML qui permet d'organiser des métadonnées variées se rapportant à un objet numérique complexe (revues numériques, sites Web...). Les **spécifications** de METS sont exprimées à l'aide d'un **schéma** XML. A cette adresse
  - (http://www.loc.gov/standards/mets/), on trouvera de la documentation, le schéma XML, des exemples...

Les métadonnées des thèses électroniques françaises TEF, Groupe AFNOR CG46/CN357/GE5, seconde Édition, Mars 2006

METS (2)

METS se décompose en 7 sections comprenant chacune différents types de métadonnées :

1. ***metsHdr*** donne des informations sur le fichier METS lui-même (date de création, créateur, identifiant…
2. ***dmdSec*** comprend l'information bibliographique
3. ***amdSec*** comprend des métadonnées de gestion du document numérique décrit, divisées en 4 sous-sections : ***sourceMD*** - ***techMD*** - ***digiprovMD*** - ***rightsMD***
4. ***fileSec*** fait l'inventaire de l'ensemble des fichiers qui composent le document (master, ocr, table des matières…), avec leurs caractéristiques techniques: taille, empreinte, emplacement du fichier… ;
5. ***structMap* (seule section obligatoire)** : structure du document (niveaux de granularité) :
6. ***structLink*** : liens structurels entre différents éléments (hyperliens…)
7. ***behaviorSec*** (fichiers associés à des programmes de lecture).

OAIS (2010)

- **Information de représentation** (*Representation Information*) : information qui relie un Objet-données à des concepts plus explicites.
  - Par exemple, l'Information de représentation de la séquence de bits constituant un fichier FITS pourra être la norme FITS qui définit le format avec un dictionnaire donnant le sens, dans ce fichier, des mots-clés qui ne figurent pas dans la norme.

- **Information de pérennisation (***Preservation Description Information - PDI***)** : information nécessaire à une bonne conservation du Contenu d'information, et qui peut être décomposée en Informations de provenance, d'identification, d'intégrité, de contexte et de droits d'accès.

XML Formatted Data Unit (XFDU) - Structure and Construction Rule

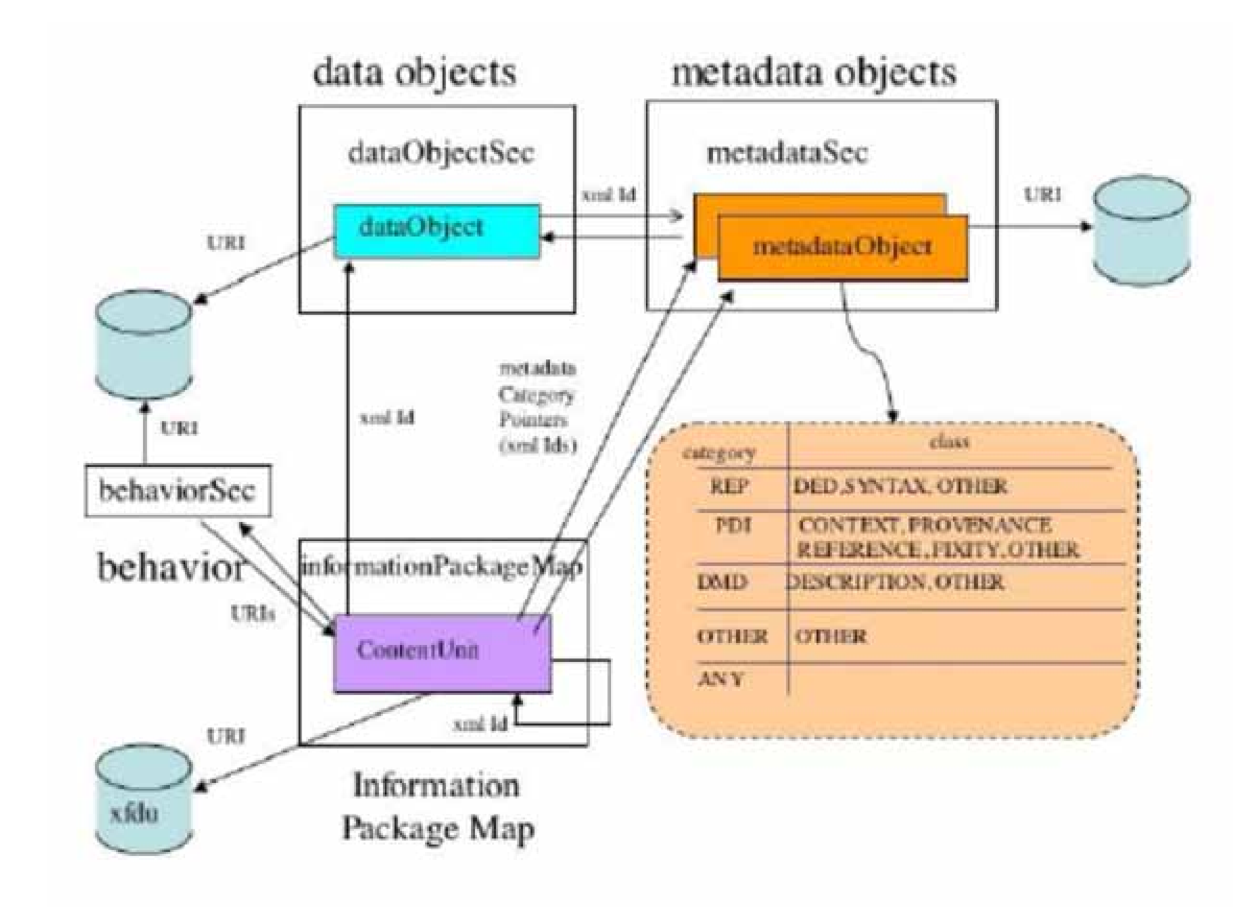


Figure 2: XFDU Manifest Logical View

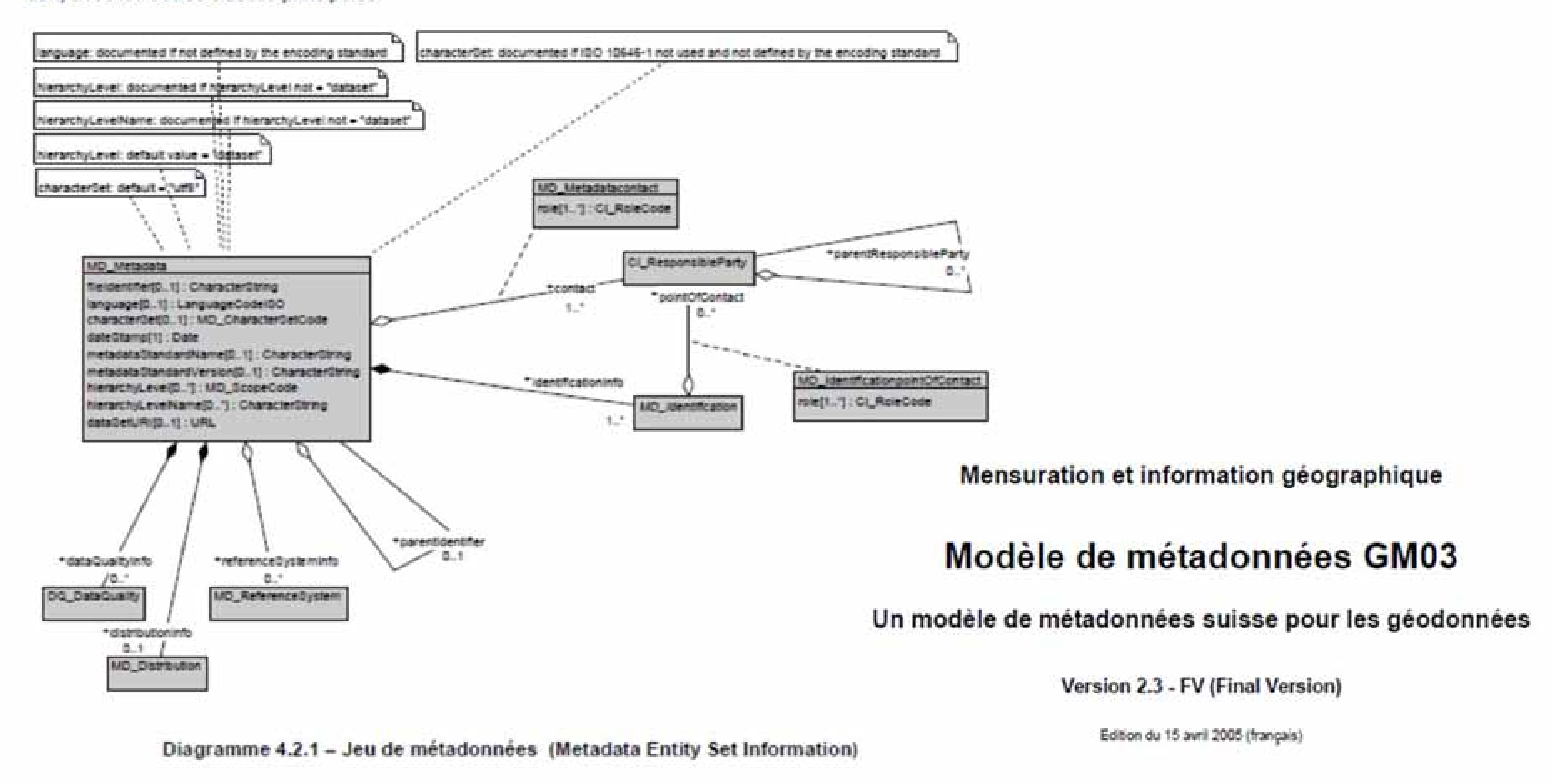
CCSDS

XFDU est un **format** de type « capsule » qui permet de créer l'enveloppe du paquet d'information. Il est très proche de METS, dont il reprend les mécanismes

XFDU est un **standard** d'empaquetage comme METS

XFDU est un **schéma** de données complexes…

Modèle GM03 – Jeu de données géographiques



4.2.1 Jeu de métadonnées (Metadata Entity Set Information)

Jeu de métadonnées comportant une information importante relative aux métadonnées, en relation (agrégation) avec les autres classes principales.

Diagramme 4.2.1 – Jeu de métadonnées (Metadata Entity Set Information)

Mensuration et information géographique

Modèle de métadonnées GM03

Un modèle de métadonnées suisse pour les géodonnées

Version 2.3 - FV (Final Version)

Edition du 15 avril 2005 (français)

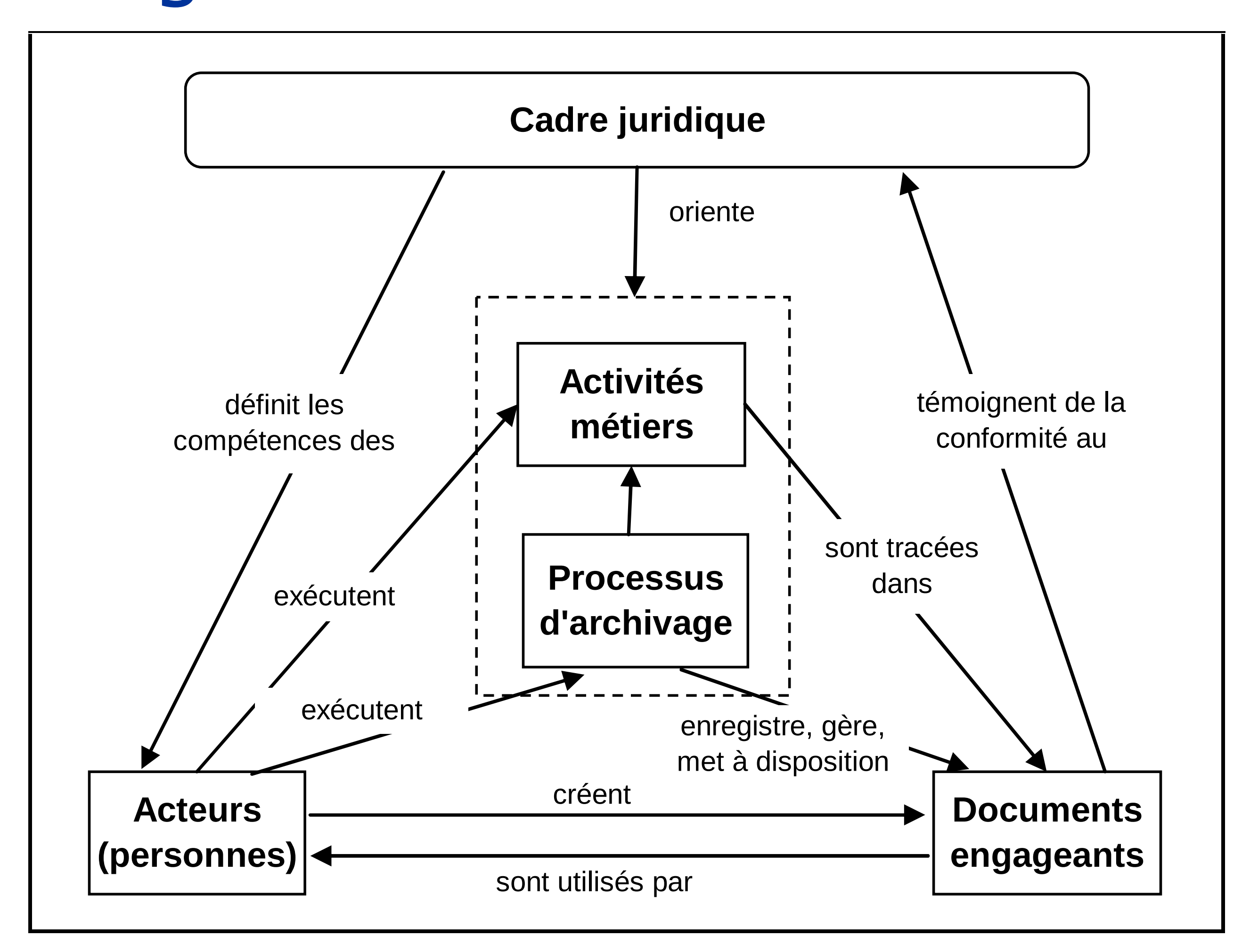
Métadonnées gouvernementales du Québec

3 outils documentaires pour les documents du Gouvernement

1. Les profils de métadonnées gouvernementaux qui s'appuie sur le Dublin Core (ISO 15836)
   - besoins de repérage de la documentation et des bibliothèques
   - **ajout d'éléments pour satisfaire aux besoins administratifs, de transactions, de tenue de dossier et de valeur probante**
2. Le *calendrier de conservation*
3. Le Thésaurus de l'activité gouvernementale et les *schèmes de classification*

http://www.banq.qc.ca/services/archivistique_ged/crggid_outils/profils_metadonnees/index.html

Norme ISO 23081 sur les métadonnées d'archivage



Trad MAC 2009

ISO 23081 et ICA-Req/ISO 16175

Métadonnées d'archivage / Records management metadata	Métadonnées qui identifient, authentifient et contextualisent les documents engageants ainsi que les personnes, processus et systèmes de création, gestion, conservation et utilisation afférents, et les règles associées. Voir aussi Métadonnées Source: ISO 23081, Part 1, Clause 4.

ICA-Req, Module 2 - 3.4.7 - Métadonnées d'archivage

Il est nécessaire de fournir des métadonnées sur les processus de gestion du document engageant archivé, y compris son sort final : toute altération, tout lien et toute utilisation du document doit être tracé(e) dans le temps de manière incontestable pour **garantir l'intégrité et l'authenticité du document**.

Même si le document engageant est figé et inviolable, les métadonnées d'archivage continueront à s'accroître durant toute sa vie administrative. Elles doivent être liées de manière permanente au document, pour garantir son authenticité, son intégrité et sa fiabilité.

ICA-Req – Principes et exigences fonctionnelles pour l'archivage dans un environnement électronique , 2008

Le modèle de métadonnées de MoReq2 (1)

- Le modèle de métadonnées de MoReq2 a deux objectifs assez proches :
  - définir les métadonnées nécessaires à **l'échange** des documents archivés entre systèmes d'archivage sans perte de fonctionnalités indispensables pour MoReq2, en tenant compte des exceptions ci-dessous ;
  - les définir avec suffisamment de précision pour produire, développer et utiliser un schéma **XML** pour MoReq2

- L'historique des événements est utilisé pour stocker les anciennes valeurs de métadonnées modifiées

MoReq2 « Exigences types pour la maîtrise de l'archivage électronique »

Le modèle de métadonnées de MoReq2 (2)

Exemple de métadonnée	**Exemples de valeurs possibles**
Titre	Demande d'autorisation XXX ou Rapport sur l'accident YYY ou Plainte au sujet de ZZZ
Identifiant	N1128A ou 3F2504E0-4F89-11D3-9A0C-0305E82C3301 ou 7QDBkvCA1+B9K/U0vrQx1A
Mots-clés	Document archivé, Archives historiques, Information ou Voyage en avion ou Activités de loisirs, sport, compétition, javelot

Le modèle de métadonnées de MoReq2 (3)

- Le modèle de métadonnées se présente comme un **jeu** de métadonnées minimal. Il s'agit des métadonnées qu'un SAE doit pouvoir exporter, importer et traiter. Chaque métadonnée correspond à ce qu'on appelait naguère un « champ », utilisé pour contenir la valeur de la métadonnée. Le tableau ci-après donne des exemples de métadonnées et de valeurs possibles.

MoReq2 « Exigences types pour la maîtrise de l'archivage électronique »

Le modèle de métadonnées de MoReq2 (4)

Caractéristiques des métadonnées Exemple - M065 Description.date

	Statut	Obligatoire	**Occurrence**	Unique	
Définition	Date du document archivé (telle qu'elle figure dans le corps du document).				
Entité(s)	série	dossier	sous-dossier	volume	document X
Alimentation	Capture automatique par le SAE si possible. Sinon saisie manuelle lors de la capture.				
Source	Document archivé ou utilisateur.				
Valeur par défaut	Valeur la plus récente utilisée pour cet élément (si elle ne figure pas dans le document).				
Héritage	Néant.				
Conditions d'utilisation	Non modifiable.				
Commentaire	Capture automatique par le SAE pour les courriels et pour les documents bureautiques élaborés avec le modèle approprié. Pour les courriels, la valeur peut être extraite du champ « Date orig. » de RFC 2822. Saisie manuelle en cas d'utilisation d'un gabarit incompatible.				
Exigences	6.1.18				

Les métadonnées dans MoReq10

- **Metadata element** (noun), An item of metadata described by a metadata element definition that has zero or more **values** given to it by users and by the MCRS*.

- **Metadata element definition** (entity) A definition of a metadata element indicates, among other **properties**, its:
  - Title – the name of the metadata element;
  - Datatype – what type of metadata it may contain;
  - Cardinality – how many values it may have; as well as
  - Whether these values may be changed by a user.

*** MCRS** = **MoReq** Compliant Records System

http://moreq2010.eu/.

Exemples de formats concurrents

- **DTD EAD** (Encoded Archival Description, définition de type de document),
- Etats Unis, 1998
- permettant de structurer en XML (*eXtensible Markup Language*) les instruments de recherche de type archivistique :

- **ISAD(G)** : Norme générale et internationale de description archivistique, 2e édition, ICA, 2000.
  - Lignes directrices pour la description archivistique
  - 26 éléments de description à combiner pour constituer la description d'une entité archivistique

EAD et ISAD(G) sont compatibles

La jungle…

modèle
DTD EAD
standard
ISO 19115
obligatoires
ISO 14721
administratives
ISAD(G
OAIS
descriptives
schéma
ISO 15836
contextuelles
FORMAT
modifiables
METS
norme
ICA-Req
structurelles
de pérennisation
LOMFR
interrogeables
de publication
Dublin Core
RFD
MoReq2
internes
dictionnaire
évolutives
de préservation
techniques
facultatives
empaquetage
DE CONTENU
sémantique
XFDU
d'environnement
statiques
registre
externes
ISO 16175
ISO 23081
profil
EAD

Reformulation, cartographie et itinéraires

I – Reformulation

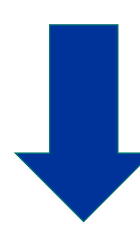
- Les objets « métadatés »
- Les métadonnées
- Les normes

Impact du numérique

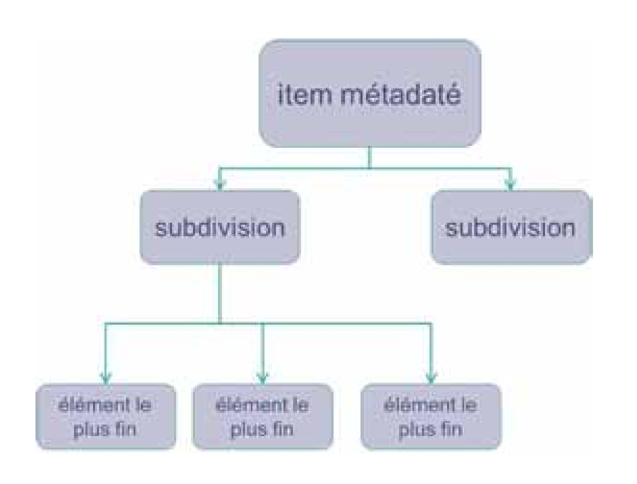
document de plus
en plus complexe



métadonnées de plus
en plus complexes



question de la granularité des
objets « métadatés »



- Normes dédiées aux métadonnées
  - stabilisation autour des grands modèles existants
  - enrichissement, perfectionnement

- Normes incluant des préconisations sur les métadonnées

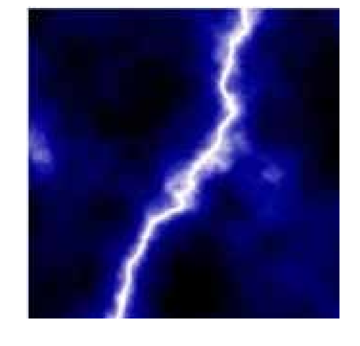
Exemples



From: Laurence PATTIER
To: xavier.dubreuil@torpica.com
Cc: Gaston Marchand
Sent: Friday, May 07, 2010 2:39 PM
Subject: Re: Projet 1066

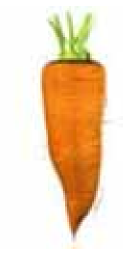
OK, 22.

LP
DCLS
Tommaco and Co
01 48 39 39 48









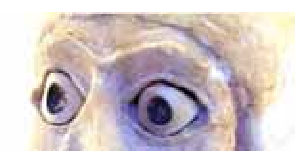








Reformulation (1)-Définition des métadonnées

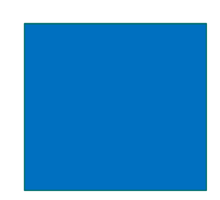


- Données
- **Associées** à des données
  - soit lors de la production des données pour compléter leurs sens et en permettre l'interprétation
  - soit au cours du cycle de vie de l'objet d'information ainsi constitué pour en faciliter l'usage et l'exploitation
- **Dissociables** des données initiales pour un traitement spécifique (ressource ou trace)

MAC2011

Reformulation (2)-Finalité des métadonnées

1. Avoir un document complet (autoportant, tracé, interprétable, exploitable)



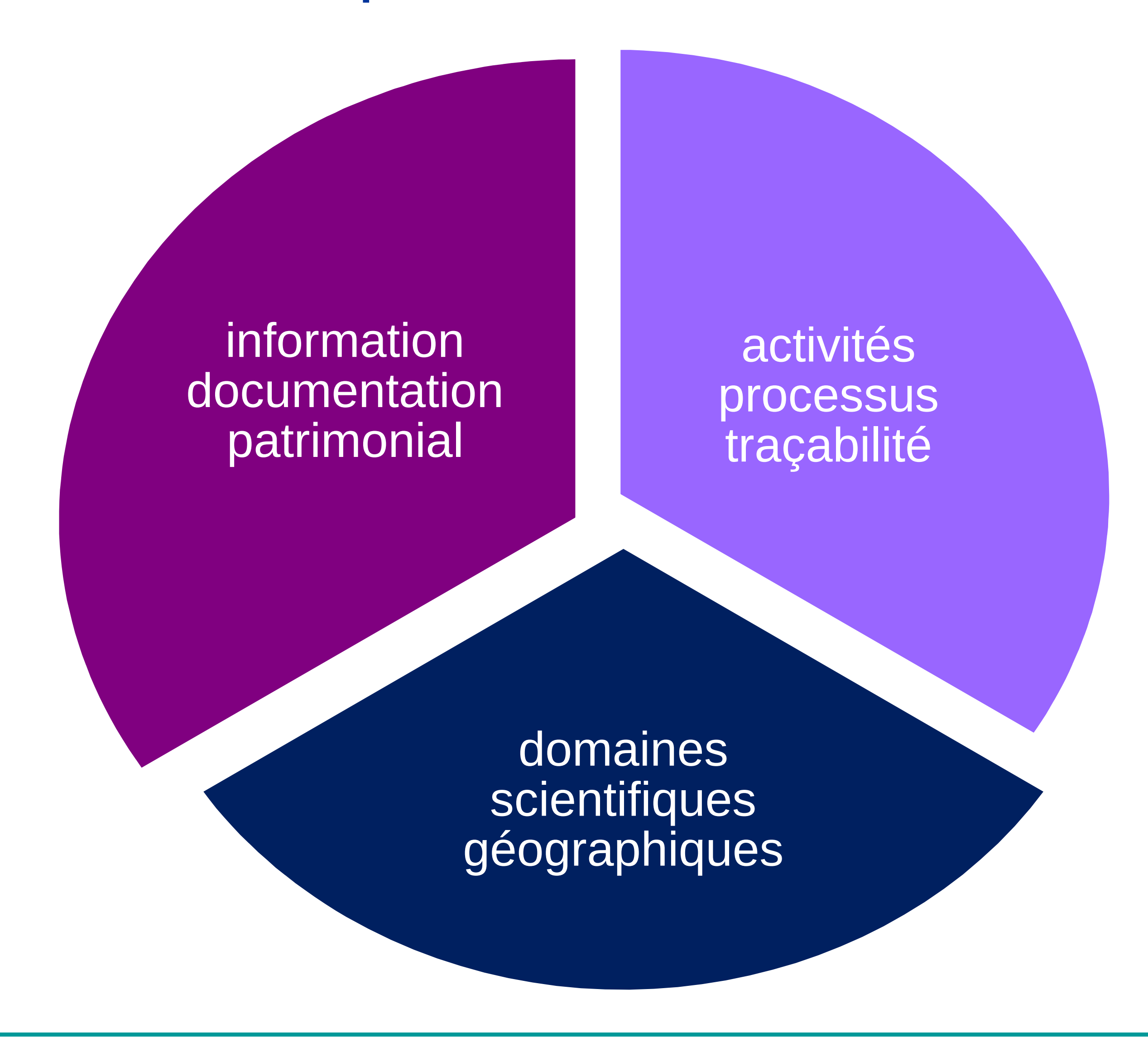
2. Pouvoir l'utiliser de manière fiable pendant toute sa vie

Reformulation (3)- Relation métadonnées/données

- Ce qui fait partie intégrante du document au moment de sa production, inclus dans le fichier ou non

- Ce qui participe de son usage: maintenance, exploitation, enrichissement, sans influer sur la valeur initiale du document

Trois domaines pour l'information



Trois domaines pour les normes de métadonnées

Dublin Core
(ISO 15836)

information
documentation
patrimonial

ISO 23081

activités
processus
traçabilité

domaines
scientifiques
géographiques

ISO 19115

II – Cartographie et itinéraires

- Deux grands types de métadonnées
- Articulation des textes normatifs
- Cartographie des métadonnées
- Conclusion

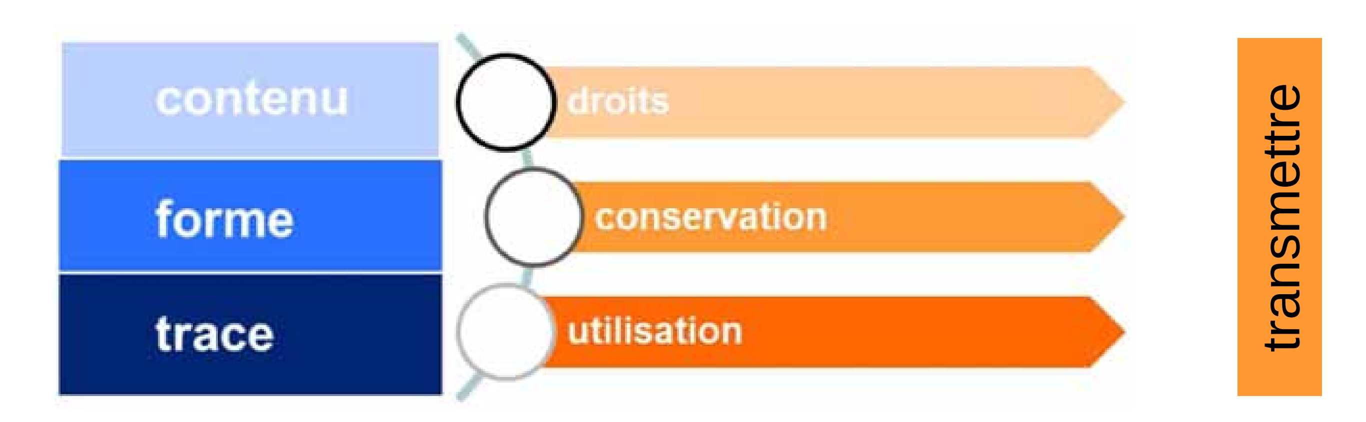
Deux grands types de métadonnées



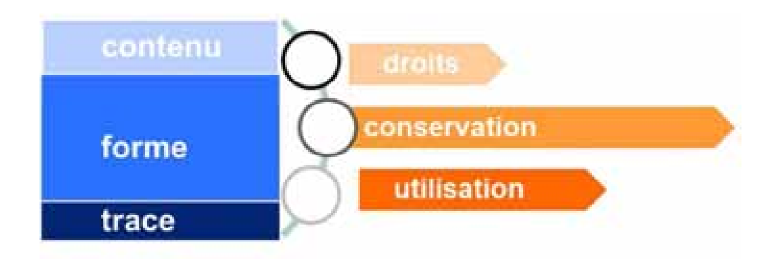
Métadonnées sub-document
préexistantes (modèle, contexte) et cocommitantes (étapes de production): message, support, responsabilité

Métadonnées post-document
extérieures ou liés au cycle de vie après la production: conservation, usage, enrichissement

supporter

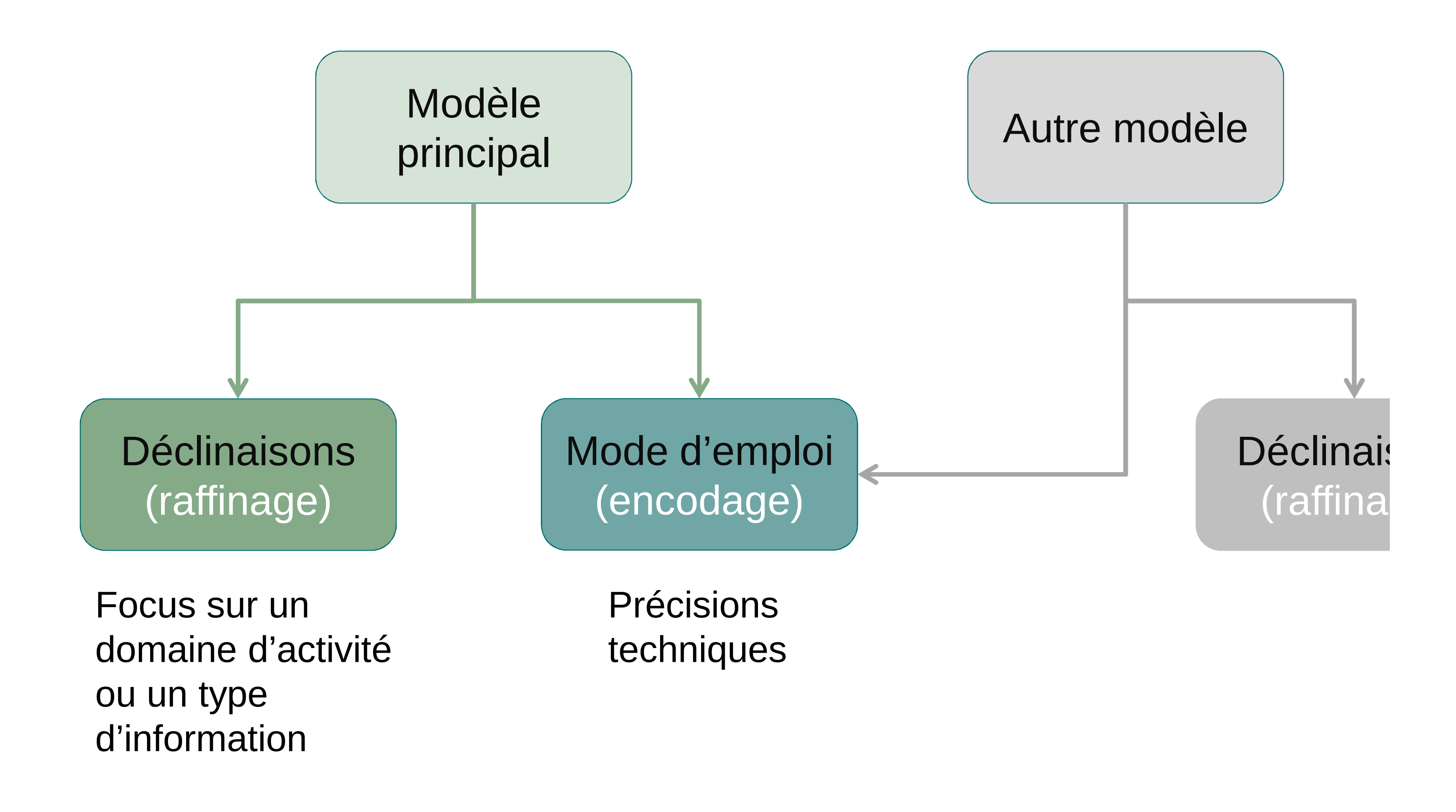


variantes des proportions, ex:

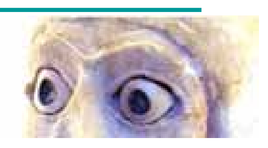


MAC2011

Articulation des documents normatifs



La grande ville des métadonnées: plan d'orientation



ISO 15836 (Dublin Core)

ISAD-G

ISO 23081

MoReq

ICA-Req

LOMFR

EAD

METS

Québec

RFD

PREMIS

XFDU

OAIS

ISO 19139

ISO 19115

GM03

MAC2011

Face à la jungle

- Caractéristiques de la jungle
  - variété des espèces
  - croissance
  - enchevêtrement



- Objectifs
  - s'y repérer
  - supprimer les éléments qui empêchent d'avancer

Deux questions

- où suis-je?
  - domaine de l’information
    - expression
    - observation
    - gestion
- où vais-je?
  - rôle des métadonnées
    - produire
    - conserver
    - exploiter

Quelques réponses...

- Tout est métadonnées
  - *le mot perd de l'intérêt…*
- Le raffinement et l'encodage
  - *le fond (l'organisation des données) et la forme (interopérabilité à tous les étages)*
- Économie des métadonnées
  - *Conception/production versus exploitation*

Alors, la « jungle »…?

modèle
DTD EAD
standard
ISO 19115
obligatoires
ISO 14721
administratives
OAIS
ISAD(G
descriptives
schéma
ISO 15836
contextuelles
FORMAT
modifiables
ICA-Req
METS
norme
structurelles
de pérennisation
LOMFR
interrogeables
de publication
MoReq2
Dublin Core
RFD
internes
dictionnaire
évolutives
de préservation
techniques
facultatives
empaquetage
DE CONTENU
sémantique
XFDU
d'environnement
statiques
registre
externes
ISO 16175
ISO 23081
profil
EAD



oui, mais une gentille jungle …

pour poursuivre la discussion

http://pin.association-aristote.fr

marie-anne.chabin@archive17.fr
www.marieannechabin.fr